Kenyon College

Digital Kenyon: Research, Scholarship, and Creative Exchange

IPHS 300: Artificial Intelligence for the Humanities: Text, Image, and Sound

Digital Humanities

Fall 2019

Analyzing popular music using Spotify’s Machine Learning Audio Features

Noah Amsterdam
Kenyon College, amsterdam1@kenyon.edu

Follow this and additional works at: https://digital.kenyon.edu/dh_iphs_ai

Part of the Digital Humanities Commons

Recommended Citation

Amsterdam, Noah, "Analyzing popular music using Spotify’s Machine Learning Audio Features" (2019). *IPHS 300: Artificial Intelligence for the Humanities: Text, Image, and Sound.* Paper 14.
https://digital.kenyon.edu/dh_iphs_ai/14

This Poster is brought to you for free and open access by the Digital Humanities at Digital Kenyon: Research, Scholarship, and Creative Exchange. It has been accepted for inclusion in IPHS 300: Artificial Intelligence for the Humanities: Text, Image, and Sound by an authorized administrator of Digital Kenyon: Research, Scholarship, and Creative Exchange. For more information, please contact noltj@kenyon.edu.

Analyzing popular music using Spotify's Machine Learning Audio Features

Artificial Intelligence for the Humanities

Noah Amsterdam, Kenyon College Class of 2022

Introduction

The modern ability to stream music using services such as Spotify, Pandora, and Apple music has revolutionized how it is consumed. There is now more music accessible at our fingertips than ever before in history. As someone who has played an instrument their whole life and with some music theory knowledge, I always strive to better understand what I'm listening to. This motivated me to dive into finding out how music is quantified, specifically on Spotify, and what makes up people's musical taste.

I will be approaching music tastes from 3 angles: Spotify, music critics, and my own analysis and knowledge. To do so I am using a dataset that includes the top 100 most popular songs from 2018, off the playlist made by Spotify. This represents what Spotify and the public consider to be the best of the year. I will be comparing this to two separate top 10 lists of the '10 worst songs of 2018' to get the critics perspective. While some of the songs were shared between both lists, of the 16 I use, interestingly 6 of them are also in the top 100 playlist. In order to attempt to understand this anomaly, I will be comparing songs based off the variables Spotify uses to quantify their music (I get into the specifics below) and analyzing the clustering and outliers I see when graphing them. I also hope to bring some of my music theory knowledge, as well as a more contextual approach into understanding why popular music can be simultaneously very successful, while also being considered poor quality.

About the Dataset

To test how Spotify breaks down music into quantifiable variables, I used a Kaggle notebook containing the top 100 songs on Spotify from 2018 charts off Spotify as my dataset. I then added 16 songs "bad" songs, one from TIME Magazine and one from Spin, again with 6 songs present in both the lists and the playlist.

On their website for developers, Spotify lists the features it uses with songs on their services, giving brief explanations for what goes into each one. This includes: Key, mode (major/minor), tempo, time signature, instrumentalness, acousticness, liveness, loudness, speechiness, and what I feel to be the 3 most important: danceability, energy, and valence. I picked these because these are the most specific to Spotify and involve the most emotion. Most of these variables are scored from 0.0-1.0, a score closer to 1 being higher.

Danceability is determined by a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.

Energy is a representation of intensity and activity, usually fast, loud, and noisy. The more specific features include "dynamic range, perceived loudness, timbre, onset rate, and general entropy."

Valence is a measure of the track's musical positivity, with a higher value sounding more positive.

The Kaggle notebooks I used included visualizations of the relationship between these features. They then use 'nearest-neighbor' to measure how similar tracks are to each other.

In using this dataset, I looked to find which of these features, if any, were particularly important in determining a song's success, negative reception, or both.

Methodology

I chose to work using the coding language R, as this what the Kaggle notebooks I found used. I already had the top 100 songs in a csv file from the notebook I was using, so I simply needed to add additional song data into the list. I found the song information using Spotify's developer website which allows access to the variables that Spotify quantifies each song with.

In order to look for correlations and outliers I used a Plotly that would take multiple variables into the x-axis, using dimensionality reduction via PCA to reduce it to a 2D plot. This then put songs against a y-axis. The songs were then displayed to be analyzed by the nearest neighbor. I also had to switch to a Jupyter notebook to run the Plotly with my dataset.

Once I looked for the bad songs amongst the plotted data I noted roughly where they were in relation to the axis. The first case was the preexisting formation comparing how danceable a song was on the x-axis to how loud it was on the y-axis. Songs with higher values on the x-axis were less danceable than those on the left. Songs high on the y-axis were louder with more noise.

After taking note of the songs, I switched the variables in the x-axis and y-axis. I repeated this step looking for any clustering. For my next orientation I wanted to continue to focus on danceability but use what I felt would be a more important y-axis variable by using valence. The more positive a song was the lower it would be on the y-axis. The more danceable and energetic a song was, the further to the left it would be.

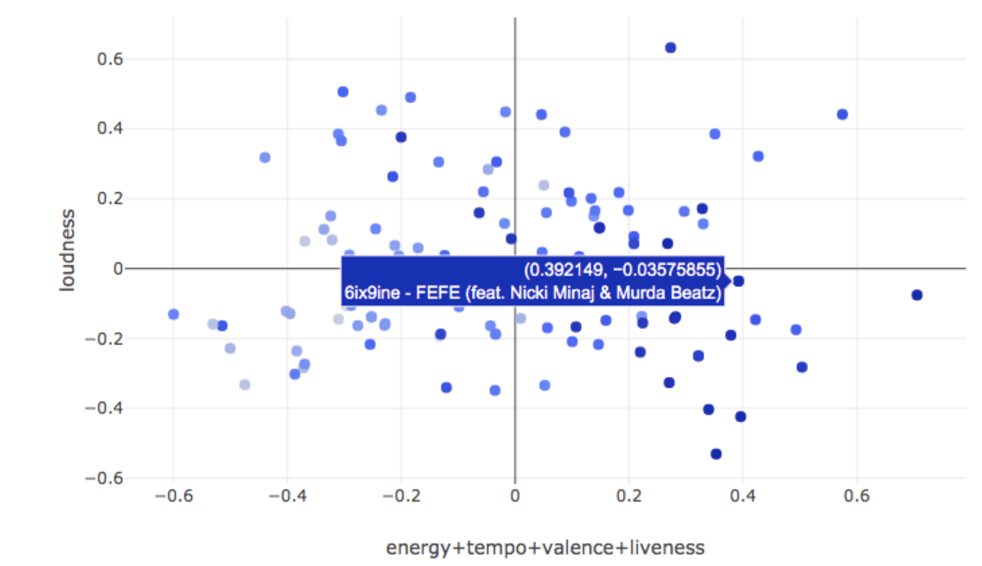
After checking where the 16 songs moved, I then continued to strip the x-axis of variables, just using energy and tempo on the x-axis against valence in the y-axix.

Finally I switched to tsne for additional dimensionality reduction, using all of the variables except the simpler ones such as Key, mode, and time signature. These were all then compared to valence. My aim in doing so would be to see if there were more overarching trends when all the variables were combined under one axis. In this graph, songs that are more positive will be lower, and the further right the song is, the lower its score of the features mentioned.

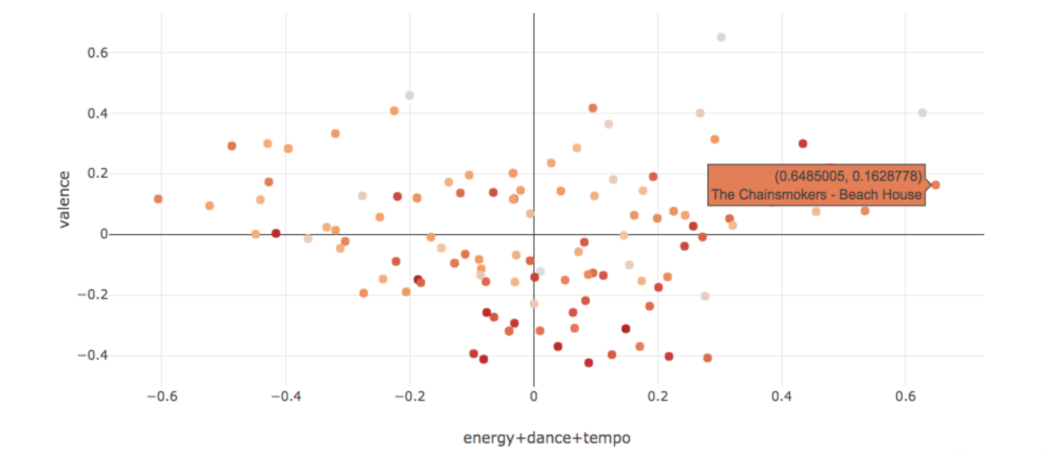
Fall	Eminem
Lover, Leaver	Greta Van Fleet
I Love It (& Lil Pump)	Kanye West
Taki Taki (with Selena Gomez, Ozuna & Cardi B)	DJ Snake
Let Me	ZAYN
Girls Like You (feat. Cardi B)	Maroon 5
The Hard Stuff	Justin Timberlake
B*TCH I'M BELLA THORNE	Bella Thorne
FEFE (feat. Nicki Minaj & Murda Beatz)	6ix9ine
Psycho (feat. Ty Dolla $ign)	Post Malone
Fuh You	Paul McCartney
Live or Die	Noah Cyrus
Lights Down Low	MAX
Lucid Dreams	Juice WRLD
All The Stars (with SZA)	Kendrick Lamar
Beach House	The Chainsmokers

Segment of dataset containing 16 of the worst songs of 2018

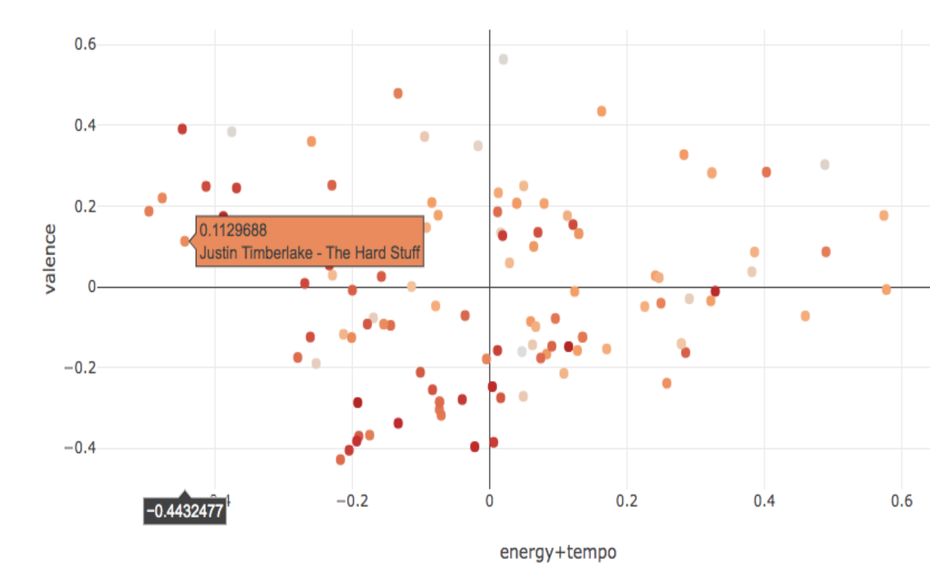
Results



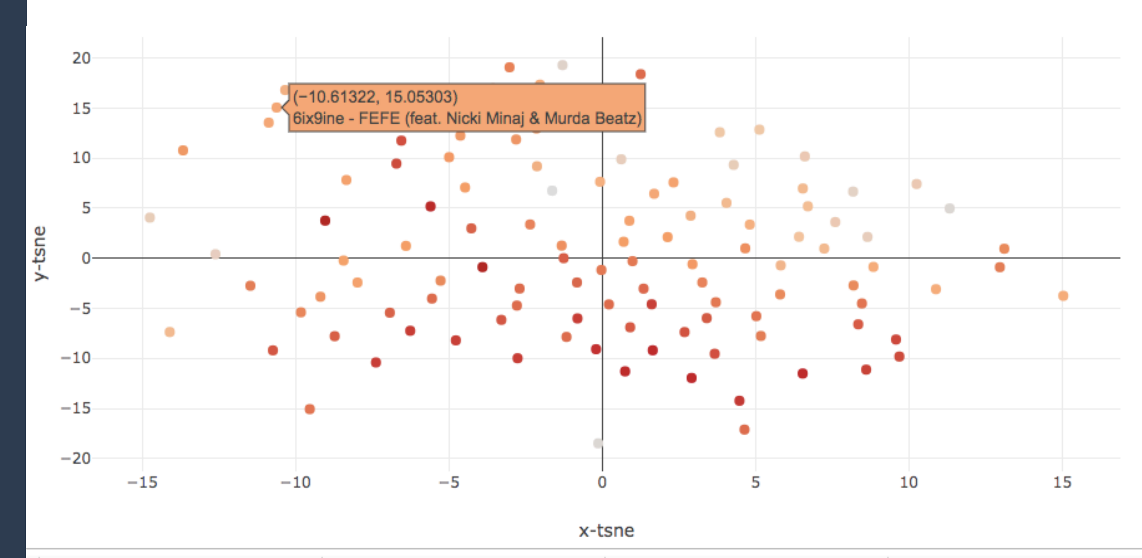
Case 1 (preset): Energy, tempo, valence, and liveness against loudness



Case 2: Energy, danceability, and tempo against valence



Case 3: Just Energy and tempo against valence



Case 3: All variables against valence (using tsne)

Conclusion

Throughout my analysis, it was clear that finding correlation would be difficult with such a small sample size of songs ranging in genre. I found that with case 1, the song's placements all made sense to me musically, but gave little insight into similarities between songs other than than whether they were loud, soft, danceable, or not. The overall scatterplot showed a relatively even split amongst the four sections, however the points do seem to favor songs that are danceable. Unfortunately, this formation says little about why some of these songs were reviewed badly.

Similar results were seen when I changed the y-axis to valence. Whether a song is positive or negative seemed to have little bearing on how good it was musically.

The closest indicator of a grouping was seen in the final 'tsne' case where I used all variables against valence. While valence was not a factor, I did notice that many of the bad songs were on the right except a few outliers which were on the far left. While not conclusive, overall it did show that with some of the negative songs they shared some characteristics with their low combination of danceability, energy, etc. score.

What these outcomes really show is that the popularity of a song is dependent on a much wider scope of factors than those that can be quantified musically. Many of the complaints in the reviews were embedded too heavily in context, involving the artist's previous work, the creativity and originality in lyricism, and the innovation in sound.

From a music theory perspective, finding that there is not yet telling evidence is understandable considering most pop songs tend to stick to a similar structural formula. Ultimately it is quite clear that with popular music the context involving a song is sometimes more important than what's happening musically in the song itself. The lyrics, artist's image, and cultural significance all have a greater impact on success than the features Spotify uses.

Future Analysis

Using Spotify's musical features to understand the anomaly of badly reviewed music being so popular, ultimately failed. However, I do see a lot of potential with other approaches. In using one relatively similar, I would simply change the dataset to include only songs within the same genre or by the same artist. This would ensure most of the songs are musically similar enough to then understand why some were hits, flops or both.

With my main issue being a lack of context concerning the artist, another possibility involves tackling the subject of lyrics. This could be done using sentiment analysis to more effectively group emotions and themes. This could be used to see how original the ideas are in a song, which could possibly be an indicator of how receptive critics would be toward it.

Ultimately, finding a metric to measure how good something is the difficult obstacle to overcome when music is so subjective. Finding a way to objectively do so could lead to massive progress in understanding why a song becomes successful.

References

Aeryan. (2017, December 11). Spotify music analysis. Retrieved December 19, 2019, from https://www.kaggle.com/aeryan/spotify-music-analysis.

Bogacz, M. (2019, December 6). What did we listen to on Spotify in 2018? . Retrieved December 20, 2019, from https://www.kaggle.com/michau96/what-did-we-listen-to-on-spotify-in-2018.

Get Audio Features for a Track. (n.d.). Retrieved December 20, 2019, from https://developer.spotify.com/console/get-audio-features-track/?id=06AKEBrKUckW0KREUWRnvT.

Get Audio Features for a Track. (n.d.). Retrieved December 20, 2019, from https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/.

Staff, S. P. I. N. (2018, December 28). The Worst Songs We Heard in 2018. Retrieved December 20, 2019, from https://www.spin.com/2018/12/worst-songs-of-2018/.

Staff, T. I. M. E. (2018, December 13). The 10 Worst Songs of 2018. Retrieved December 20, 2019, from https://time.com/5475647/worst-songs-2018/.

Acknowledgements

Thank you to Professor Chun for all the help on this project.